

Weekly Report

09/29/2014 - 10/12/2014

Jing XIA

October 12, 2014

1 Summary

This week I mainly focused on the rank visualization project. I followed the idea described in the last weekly report and implemented the algorithms. For the visualization part, I consulted Victor Chen and Cheryl Qian. They provided several good suggestions and design schemes.

2 Projects

2.1 Project 1 - Rank Visualization

2.1.1 Similarity and Clustering

Last weekly report mentioned a paper *Online Discovery of Group Level Events in Time Series* [1], which discussed group level events (group formation and group disbanding) in time series with AutoDBSCAN algorithm. I described my concern about applying the algorithm in time series clustering. The email I sent to the author was replied. The author says “The major issue of using DBSCAN in time series is that DBSCAN in general cannot work well in high dimensional data but the number of features in time series is very high. Here, we assume that the temporal autocorrelation of time series is strong and therefore the intrinsic dimension is not very high”. Thus, AutoDBSCAN may not be suitable for our data.

I’ve implemented the algorithm described in the paper *Curve Matching, Time Warping, and Light Fields: New Algorithms for Computing Similarity between Curve* [2]. The steiner point interpolation is not suitable for our data, because the two dimensions in rank time series are time and rank order, which are not compatible to each other. It’s not like normal curves which can take two directions equivalent. After testing the algorithm with some sample rank time series, I removed the steiner point part. Thus the algorithm is somewhat similar with our original ones, except replacing line path with manifold path (see Figure 1).

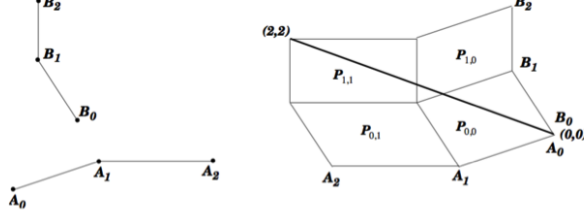


Figure 1: Polygonal chain and the corresponding manifold. The similarity of the two chains can be equivalent to distance between $(0, 0)$ and $(2, 2)$.

2.1.2 Evaluation

I've also implemented the evaluation function in *Online Discovery of Group Level Events in Time Series*[1]. For the entropy of a cluster χ of m time series (x_1, x_2, \dots, x_m) at time t

$$S(\chi, t) = - \sum_{j=1}^m \frac{1}{m} \log \left(\sum_{i=1}^m \exp(-d(x_i, x_j, t)) \frac{1}{m} \right) \quad (1)$$

The way we use this function is still in consideration. Whether to present it with visualization or to just evaluate in the paper is to be decided.

2.1.3 Visualization

I want to acknowledge Dr. Victor Chen and Dr. Cheryl Qian here for their graceful help in visualization design. I've met them twice and gained much great design ideas for this project. There are several issues criticized by reviewers and we've been discussing for long but cannot come up with a fair solution. In discussion with Dr. Chen and Dr. Qian, we yield the problems with glyph design tricks.

Consistency of Item We've discussed a lot in previous versions, about color and layout of ranked items. Reviewers described their confusion towards item encoding. Now we have some *guidelines*: original rank seems better than rank difference; items should be placed corresponding to their rank orders; users should be able to locate items along time somehow; follow users' instinct in finding items. Based on these guidelines, we have the following design (see Figure 2): items are initialized in grey (Figure 2(a)) and turned to colors indicating their rank orders of the day (Figure 2(b)) after users clicks a certain day. Items in other columns are colored corresponding to the rank order of the selected day, and those not on the rank are filtered.

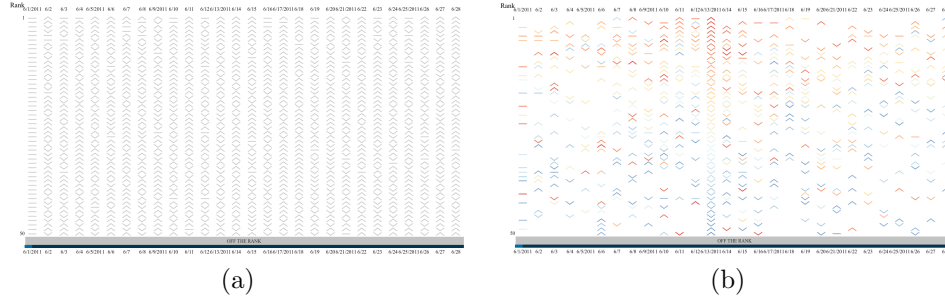


Figure 2: Items are initialized in grey (a) and turned to colors indicating their rank orders of the day (b) after users clicks a certain day. Items in other columns are colored corresponding to the rank order of the selected day, and those not on the rank are filtered.

Based on this visualization, users can locate interesting items, those of high order on the selected day but low order on other days (red items at the bottom) or those of low order on the selected day but high order on other days (blue items at the top). Also, the blank space indicates how items on the selected day occupy other columns (popularity of items).

Visual Hint In the last example, glyph design just indicates up trend, down trend and steady trend from items' last order. When highlighting the item, the glyph does not follow the curve. Thus we came up with glyph that follows the trend of its curve, namely trend from their last rank order and to the next rank order. It turns to be the Figure 3 below. Altogether there are 9 combinations from both sides of the three trends. Even without highlight the curve, users are able to indicate items' last order and next order roughly.

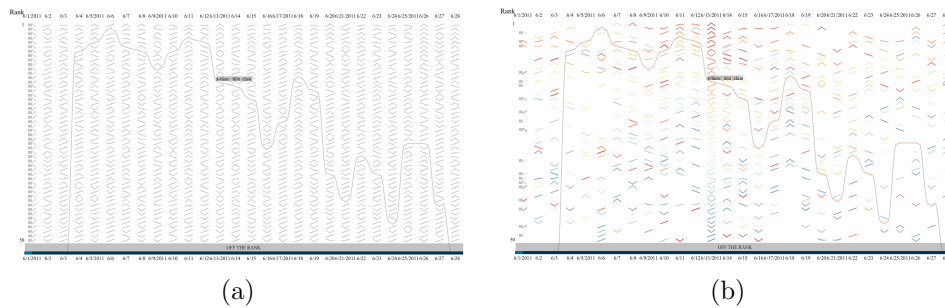


Figure 3: The glyph is designed to follow the trend of its curve.

Extended Designs - We can extend the visual hint to more accurate angle that indicates exact position of items' last order and next order.

- Items initialized with spark lines may convey more information?

- Instead of color encoding glyph, what about color encoding the area of the glyph? It makes blank space outstanding.

2.1.4 More Idea

Dr. Cheryl Qian came up with an idea of building Wikipedia ontology based on rank items of similar rank trends. Researchers usually build Wikipedia ontology based on page-to-page link, but with similar rank trends, users is able to build a stronger (weighted) Wikipedia ontology network. However, for such task, top 100 rank items are not enough for mining similar items. Say, for item “Steve Jobs” users can find his wife, his daughter and Apple’s cofounder in Top 100, but there are other items in Top 10000 maybe.

We may extend the project for *visualization of Wikipedia ontology* as a following research work, integrating the page-to-page link dataset and the page view dataset.

2.2 Project 2 - Data Inspection

Not ready to summarize yet.

3 Paper Reading

4 Miscellaneous

Reviewed the paper *Visual Analysis of Sentiment on Micro-blog*.

5 To Do List

1. Try out more visualization design schemes.
2. Complete implementation preparation of rank visualization.
3. Professor Min Chen from Oxford University is visiting VACCINE next week. Hope we can discuss the Data Inspection project then.

References

- [1] Xi C Chen, Abdullah Mueen, Vijay K Narayanan, Nikos Karampatziakis, Gagan Bansal, and Vipin Kumar. Online discovery of group level events in time series. Technical report, SIAM, 2014.
- [2] Alon Efrat, Quanfu Fan, and Suresh Venkatasubramanian. Curve matching, time warping, and light fields: New algorithms for computing similarity between curves. *Journal of Mathematical Imaging and Vision*, 27(3):203–216, 2007.